

CELEBAL TECHNOLOGIES PVT. LTD.

ASSIGNMENT

WEEK 6

Configure Self-hosted Integration Runtime to Extract Data from Local Server and Load into Azure SQL Database:

Set up a Self-hosted Integration Runtime (SHIR) in Azure Data Factory to securely connect to an on-premises server. Extract data from the local environment and load it into an Azure SQL Database to enable cloud-based analytics and processing.

STEP 1: Download and install Microsoft Integration Runtime Manager.

STEP 2: Link Azure Data Factory

The screenshot displays the Microsoft Integration Runtime Configuration Manager application window. The title bar reads "Microsoft Integration Runtime Configuration Manager". The navigation menu includes "Home", "Settings", "Diagnostics", "Update", and "Help". The "Home" tab is active, showing a green checkmark icon and the message "Self-hosted node is connected to the cloud service". Below this, the configuration details are listed: "Data Factory: week5assignment", "Integration Runtime: integrationRuntime280625", and "Node: LAPTOP-GAPN9CNC". A "Stop Service" button is present. The "Data Source Credential" section shows "Credential store: On-premises", "Credential status: In sync", and "Last backup time: N/A", with "Generate Backup" and "Import Backup" buttons. At the bottom, a status bar indicates "Connected to the cloud service (Data Factory V2)" with a refresh icon.

Data Factory:	week5assignment
Integration Runtime:	integrationRuntime280625
Node:	LAPTOP-GAPN9CNC

Stop Service

Data Source Credential ⓘ

Credential store:	On-premises
Credential status:	In sync
Last backup time:	N/A

Generate Backup Import Backup

Connected to the cloud service (Data Factory V2)

STEP 3: Create new Integration Runtime(integrationRuntime280625) in Azure Data Factory

Integration runtimes

The integration runtime (IR) is the compute infrastructure to provide the following data integration capabilities:

+ New Refresh

Filter by name

Showing 1 - 2 of 2 items

Name	Type	Sub-type
AutoResolveIntegrationRuntime	Azure	Public
integrationRuntime280625	Self-Hosted	---

Settings

Nodes Auto update Sharing Links

Install integration runtime on Windows machine or add further nodes using the Authentication Key.

Name
integrationRuntime280625

Description

Self-contained interactive authoring
☒ Disable ☐ Enable

Option 1: Express setup
[Click here to launch the express setup for this computer](#)

Option 2: Manual setup

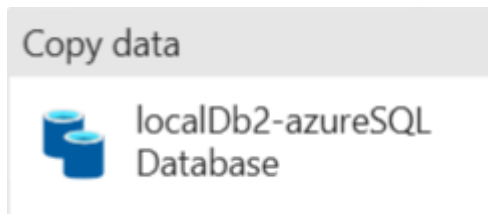
Step 1: [Download and install integration runtime](#)

Step 2: Use this key to register your integration runtime

Name	Authentication key
Key1	IR@4c687907-8091-4a13-a28f-93d5d72799f0@week5assignment@Ser
Key2	IR@4c687907-8091-4a13-a28f-93d5d72799f0@week5assignment@Ser

Apply Cancel

Tested Using: Copy data activity(localDb2-azureSQLDatabase)



DB2 (Source) Queries:

```
db2 => connect to week6

Database Connection Information

Database server      = DB2/NT64 12.1.1.0
SQL authorization ID = DEEPANSH...
Local database alias = WEEK6

db2 => CREATE TABLE employees (EmployeeID INT PRIMARY KEY NOT NULL, FirstName VARCHAR(50), LastName VARCHAR(50), Department VARCHAR(50), HireDate DATE, LastUpdated TIMESTAMP);
DB20000I The SQL command completed successfully.
db2 => INSERT INTO Employees (EmployeeID, FirstName, LastName, Department, HireDate, LastUpdated) VALUES (1, 'DEEP', 'KAUR', 'HR', '2023-01-15', CURRENT_TIMESTAMP), (2, 'ARYA', 'SHARMA', 'IT', '2022-12-05', CURRENT_TIMESTAMP), (3, 'RIYA', 'SINGH', 'Finance', '2021-06-20', CURRENT_TIMESTAMP), (4, 'ASHA', 'SAINI', 'Marketing', '2020-11-11', CURRENT_TIMESTAMP);
DB20000I The SQL command completed successfully.
db2 => SELECT * FROM Employees FETCH FIRST 1 ROWS ONLY;

EMPLOYEEID  FIRSTNAME  HIREDATE  LASTUPDATED  LASTNAME  DEPARTMENT
-----
1 DEEP      01/15/2023  2025-06-28-16.19.43.076000  KAUR      HR
1 record(s) selected.
```

Azure Data Studio (Sink) Queries:

```
SQLQuery_2 - (53) d...usaini 1

Run Cancel Disconnect Change Database: week6 Estimated Plan Enable Actual Plan Parse

1 CREATE DATABASE week6
2
3 CREATE TABLE employees (
4     EmployeeID INT PRIMARY KEY NOT NULL,
5     FirstName VARCHAR(50),
6     LastName VARCHAR(50),
7     Department VARCHAR(50),
8     HireDate DATE,
9     LastUpdated DATETIME
10 );
11
12 SELECT * FROM employees;
```

Configure FTP/SFTP Server and Create an ADF Pipeline for Data Extraction:

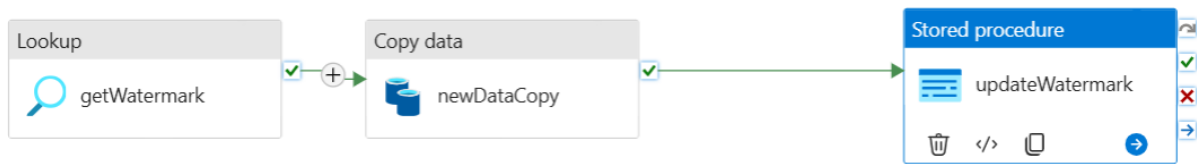
Set up access to FTP/SFTP servers and build a pipeline in Azure Data Factory to extract files or datasets from these sources. This facilitates integration with external systems or partners.

STEP 1: Log in to FTP server.

Administration interface - FileZilla Server 1.10.3			
Server Window Help			
Date/Time	Info	Type	Message
29-06-2025 19:46:16	FTP Session 227 127.0.0.1	Response	220-FileZilla Server 1.10.3
29-06-2025 19:46:16	FTP Session 227 127.0.0.1	Response	220 Please visit https://filezilla-project.org/
29-06-2025 19:46:16	FTP Session 227 127.0.0.1	Comma...	AUTH TLS
29-06-2025 19:46:16	FTP Session 227 127.0.0.1	Response	234 Using authentication type TLS.
29-06-2025 19:46:16	FTP Session 227 127.0.0.1	Comma...	USER testftp
29-06-2025 19:46:16	File-based Authenticator...	Status	Realm ftps is enabled for user testftp. Continuing authentication.
29-06-2025 19:46:16	FTP Session 227 127.0.0.1	Response	331 Please, specify the password.
29-06-2025 19:46:16	FTP Session 227 127.0.0.1	Comma...	PASS ****
29-06-2025 19:46:16	File-based Authenticator...	Status	Realm ftps is enabled for user testftp. Continuing authentication.
29-06-2025 19:46:16	FTP Session 227 127.0.0.1	Response	230 Login successful.
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Comma...	PBSZ 0
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Response	200 PBSZ=0
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Comma...	PROT P
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Response	200 Protection level set to P
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Comma...	OPTS utf8 on
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Response	202 UTF8 mode is always enabled. No need to send this command
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Comma...	PWD
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Response	257 "/" is current directory.
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Comma...	TYPE I
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Response	200 Type set to I
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Comma...	PASV
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Response	227 Entering Passive Mode (127,0,0,1,227,235)
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Comma...	REST 754974720
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Response	350 Restarting at 754974720
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Comma...	RETR cleaned_job_analytics.csv
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Response	150 Starting data transfer.
29-06-2025 19:46:16	FTP Session 227 127.0.0.1...	Response	425 Error writing to file: ECONNABORTED - Connection aborted
29-06-2025 20:55:25	Admin UI	Status	Retrieving configuration from the server...
29-06-2025 20:55:25	Admin UI	Status	Server's configuration retrieved.

STEP 2: Create FTP Linked Service (FtpServer1) on Azure Data Factory

only new or modified records are processed, optimizing performance and reducing load.



STEP 1: Lookup Activity (getWatermark)

General **Settings** User properties

Source dataset * watermarkTable Open New Preview data Learn more

First row only ☒

Use query ☐ Table ☒ Query ☐ Stored procedure

Query *

SELECT LastCopiedValue FROM Db2Watermark WHERE TableName = 'employees';

Edit

Query timeout (minutes) ⓘ

Isolation level ⓘ Select...

Partition option ⓘ ☒ None ☐ Physical partitions of table ⓘ ☐ Dynamic range ⓘ

! Please preview data to validate the partition settings.

STEP 2: Copy data Activity (newDataCopy)

General **Source** Sink Mapping Settings User properties

Source dataset * week6db2 Open New Preview data Learn more

Use query ☐ Table ☒ Query

Query

SELECT * FROM employees WHERE L...

Additional columns ⓘ + New

Pipeline expression builder



Add dynamic content below using any combination of [expressions](#), [functions](#) and [system variables](#).

```
SELECT * FROM employees
WHERE LastUpdated > '{@activity('getWatermark').output.firstRow.
    LastCopiedValue}'
```

[Clear contents](#)

Activity outputs

Parameters

System variables

Functions

Variables

Search

getWatermark

getWatermark activity output

getWatermark first row

Data of the first row

OK

Cancel

STEP 3: Stored procedure Activity (updateWatermark)

General **Settings** User properties

Linked service *

week6database

Test connection Edit New

Integration runtime *

integrationRuntime280625

Edit

Stored procedure name *

[dbo].[updateWatermark]

☒ Enter manually

Stored procedure parameters

Import New

STEP 4: Scheduled Trigger (week6ScheduledTrigger)

Edit trigger

Name *

week6ScheduledTrigger

Description

Type *

ScheduleTrigger

Start date * ⓘ

6/28/2025, 4:11:00 PM

Time zone * ⓘ

Chennai, Kolkata, Mumbai, New Delhi (UTC+5:30)

Recurrence * ⓘ

Every

1

Day(s)

Advanced recurrence options

Execute at these times ⓘ

Hours

12



Minutes

0



Schedule execution times

12:00

OK

Cancel

Azure Data Studio Queries:

```
SQLQuery_2 - (53) d...usaini) 1
Run Cancel Disconnect Change Database: week6 Estimated Plan Enable Actual Plan Parse
14 CREATE TABLE Db2Watermark (
15     TableName NVARCHAR(100),
16     LastCopiedValue DATETIME
17 );
18
19 INSERT INTO Db2Watermark (TableName, LastCopiedValue)
20 VALUES ('employees', '2025-06-28 16:19:43.077');
21
22 CREATE PROCEDURE updateWatermark
23 AS
24 BEGIN
25     UPDATE Db2Watermark
26     SET LastCopiedValue = GETDATE()
27     WHERE TableName = 'employees';
28 END;
```

Automate a Pipeline to Trigger Every Last Saturday of the Month: Configure a custom time-based trigger in Azure Data Factory to run the pipeline on the last

Saturday of each month. This supports periodic reporting and batch processing without manual intervention.

STEP 1: If Condition Activity (isLastSaturday)

General **Activities (1)** User properties

Expression ⓘ @and(equals(dayOfWeek(utcNow()), 6)

Case	Activity
True	Copy data1
False	1 Activity
	No activities

Pipeline expression builder

Add dynamic content below using any combination of [expressions](#), [functions](#) and [system variables](#).

```
@and(  
  equals(dayOfWeek(utcNow()), 6),  
  not(equals(formatDateTime(addDays(utcNow(), 7), 'MM'),  
    formatDateTime(utcNow(), 'MM')))  
)
```

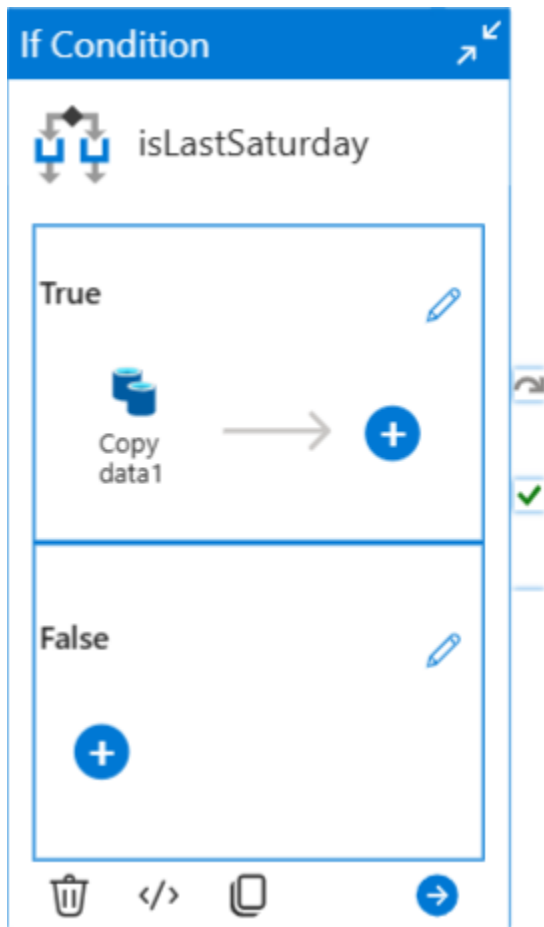
[Clear contents](#)

Activity outputs Parameters System variables Functions Variables

Search

Copy data1
Copy data1 activity output

- Copy data Activity (Copy data1) inside True Case

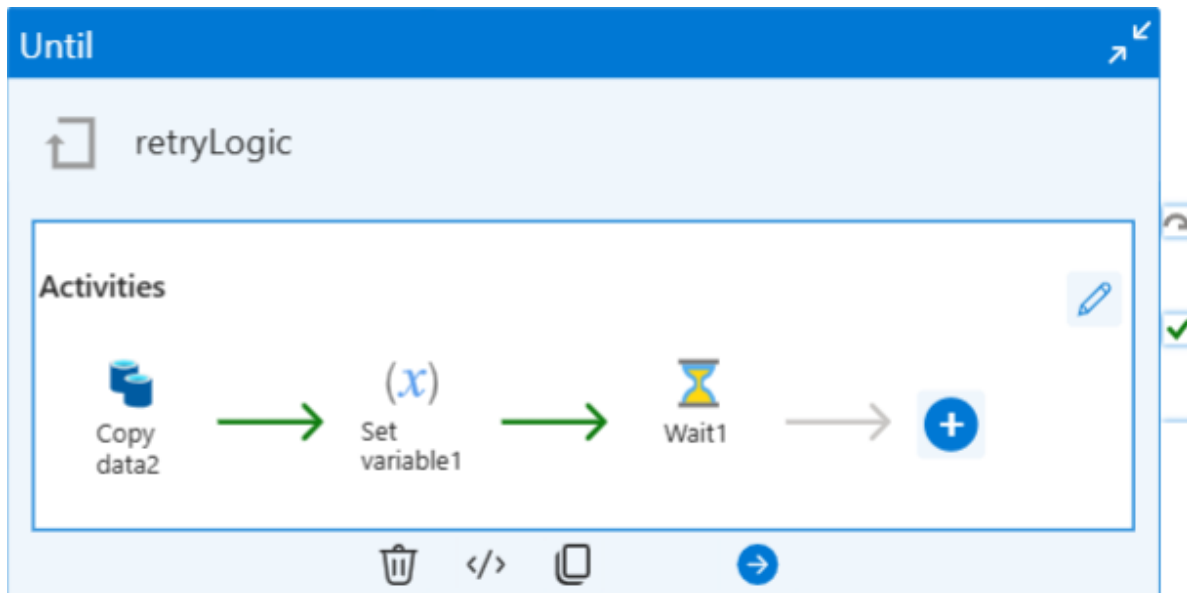


Retrieving data. Wait a few seconds and try to cut or copy again: Implement logic to handle data retrieval failures gracefully by waiting briefly (e.g., a few seconds) before retrying cut, copy, or extraction operations, improving pipeline resilience and reducing transient error impacts.

STEP 1: Until Activity (retryLogic)

General	Settings	Activities (3)	User properties
Expression	<code>@equals(variables('SuccessFlag'), true)</code>		
Timeout	0.12:00:00		

- Copy data Activity (Copy data2)



- Set variable Activity (Set variable1)

General **Settings** User properties

Variable type ☒ Pipeline variable ☐ Pipeline return value

Name * + New

Value

Pipeline expression builder



Add dynamic content below using any combination of [expressions](#), [functions](#) and [system variables](#).

```
@equals(activity('Copy data2').status, 'Succeeded')
```

[Clear contents](#)

Activity outputs

Parameters

System variables

Functions

Variables

Search

Copy data2

Copy data2 activity output

OK

Cancel

- Wait Activity (Wait1)

General

Settings

User properties

Wait time in seconds *

5